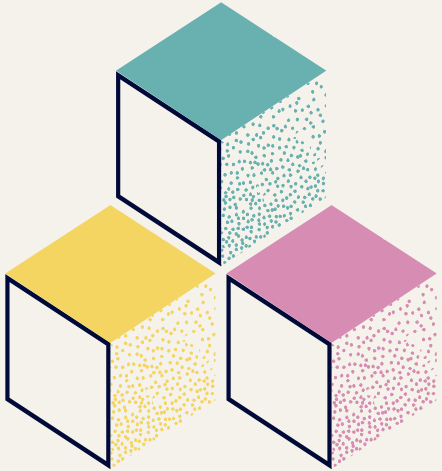


Final Year Project

First Review Presentation

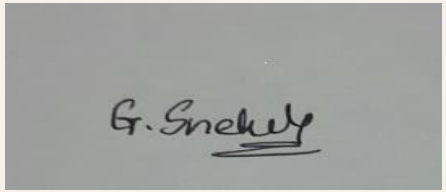
Title

Cellular Automata and Algorithmic Preprocessing to Improve Clustering

Guide

Dr. Kamalika Bhattacharjee - Computer Science and Engineering

Team 8



Ganta Sneha Rao - 106119037

Subramanian V V - 106119123

Viswonathan Manoranjan - 106119145

Introduction

Problem Statement

- A non-conventional approach towards clustering, by grouping similar data points based on the rule and cycle formation using cellular automata.
- Clustering is crucial because it determines the intrinsic grouping of the unlabeled data present.

Existing Solutions

- Exploratory search is done to find specific suitable rules for a pre-decided window size
- The given data is encoded using a 2-bit frequency encoding scheme to convert the data into binary format
- Convert each data point into a concatenated binary string
- Rules are selected randomly and applied to the data iteratively to group them into cells
- Silhouette scores are calculated and compared.

Limitations of Existing Solutions

Encoding Schema leading to loss of data

Rule Search and application

Static window size hinders expandability

Objectives

Determine the suitable hash functions for entities that need clustering to be positioned in different cells during the initial approach.

Propose a new algorithm that avoids the randomized search for rules for a particular dataset.

Package the whole process for ease of replication and development.

Provide flexibility for multiple datasets and variable window sizes.

Proposed Methodology

LIMITATION

Usage of frequency-based encoding as the encoding technique.

SOLUTION

Exploring usage of entropy-based data reduction and alternative encoding schemes such as hash functions, quantitative conversions (std dev) and Godel number encoding

LIMITATION

Exploratory search for rules and iterating through the whole set of CA rules to finalize the efficient rule increases time complexity.

SOLUTION

Find patterns correlating data to the rules to streamline the number of rules selected.

LIMITATION

Provision of rules only for a fixed window size on the encoded data.

SOLUTION

Alternate between uniform and non-uniform rule selection to facilitate variable window sizes.

Tools and Language

01

Python

For ease of understanding, functional programming, GPU compatibility and replicability of previous papers

02

Numpy and Pandas

For data handling and faster/parallel computations

03

Sklearn

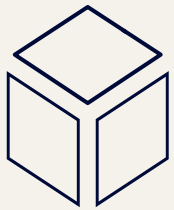
Employing K-Means clustering for baseline scores and calculating silhouette, calinski ,davies boundary scores

Timeline

Phase 1

Week 1-3

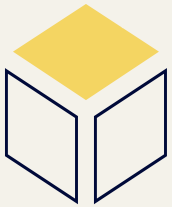
Go through the previous papers and replicate the results.



Phase 2

Week 4-8

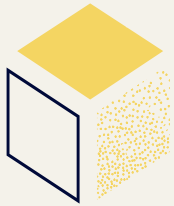
Explore various patterns and develop a suitable algorithm to cluster the data efficiently.



Phase 3

Week 9-12

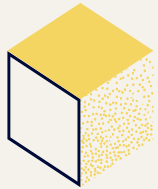
Implement and optimize the algorithm based on performance.



Phase 4

Week 13-15

Record the results and package the whole process for easy usage.



References

[1] *Sukanya Mukherjee, Kamalika Bhattacharjee, and Sukanta Das. 2020.* **Cycle Based Clustering Using Reversible Cellular Automata**. I*n Cellular Automata and Discrete Complex Systems: 26th IFIP WG 1.5 International Workshop, AUTOMATA 2020, Stockholm, Sweden, August 10–12, 2020, Proceedings. Springer-Verlag, Berlin, Heidelberg, 29–42.*

[2] *S. Mukherjee, K. Bhattacharjee and S. Das*, "**Clustering Using Cyclic Spaces of Reversible Cellular Automata**," *Complex Systems, 30(2), 2021 pp. 205–237.*

[3] *Hong He, Yonghong Tan,* "**Automatic pattern recognition of ECG signals using entropy-based adaptive dimensionality reduction and clustering**", *Applied Soft Computing, Volume 55, 2017*

[4] *Mukherjee, Sukanya & Bhattacharjee, Kamalika & Das, Sukanta. (2021)*. "**Reversible Cellular Automata: A Natural Clustering Technique. Journal of Cellular Automata"**. *16. 1-38.*

[5] *Abhishek S, Mohammed Dharwish, Amit Das, and Kamalika Bhattacharjee,* "**A Cellular Automata Based Clustering Technique for High-Dimensional Data**"*, Proceedings of the Second Asian Symposium on Cellular Automata Technology 2023, IIEST Shibpur*

Thank you!